DOCUMENT RESUME

ED 441 871 TM 031 213

AUTHOR Thomas, Sally; Smees, Rebecca
TITLE Dimensions of Secondary School Effectiveness: Comparative Analyses across Regions.
PUB DATE 2000-04-00
NOTE 20p.; For preliminary study, see ED 426 460. Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Accountability; Comparative Analysis; Educational Policy; Effective Schools Research; Foreign Countries; Geographic Location; Models; *Performance Based Assessment; *School Effectiveness; *Secondary Education
IDENTIFIERS *United Kingdom

ABSTRACT

Recent government policies in the United Kingdom implementing new systems of evaluation and accountability have highlighted the use of performance data to inform judgments about secondary schools, but these developments themselves have been informed by a relatively small number of research studies. This paper reports the findings of a study of school effectiveness from six extensive and detailed regional data sets from two sections of Lancashire, London, Jersey, Scotland, the Netherlands, and England as a whole. Objectives were to study the optimal models for measuring secondary school effectiveness across a range of outcomes in the United Kingdom and abroad, the extent of regional differences in school effectiveness results, and the underlying dimensions of school effectiveness across different regional and policy contexts. Findings show that at least four dimensions of secondary school effectiveness can be defined, in terms of different: (1) outcomes; (2) student groups; (3) pupil cohorts; and (4) curriculum stages (key stages in the British national curriculum). Regional differences appeared to exist in terms of the size and impact of school effects, and they were mirrored by differences in regional context. Effectiveness at different levels of the education system and interactions between levels need to be continually monitored in order to inform policy decisions and map out the boundaries of school effectiveness. Appendixes contain samples from the datasets, a discussion of the statistical analyses, and two tables of correlations between school effects. (Contains 4 tables and 54 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 441 871

Dimensions of Secondary School Effectiveness: Comparative Analyses Across Regions

Sally Thomas & Rebecca Smees

International School Effectiveness & Improvement Centre [ISEIC]
University of London Institute of Education

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

S. Thomas

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Paper to be presented at

American Educational Research Association
New Orleans

24-28 April 2000

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM031213

Address for correspondence:

Dr Sally Thomas
ISEIC
Institute of Education
20 Bedford Way
LONDON WC1H 0AL

Tel: +44 20 7612 6326
Fax: +44 20 7612 6344
Email: S.Thomas@ioe.ac.uk

Abstract

Recent government policy implementing new systems of evaluation and accountability have highlighted the use of performance data to inform judgements about secondary schools and stimulate school improvement. However, these developments have been informed by a relatively small number of research studies addressing the methodology of measuring school effectiveness, and often employing limited or incomplete datasets. This paper reports the findings of an ESRC funded study and employs six extensive and detailed regional datasets drawn from Lancashire, London, Jersey, Scotland, the Netherlands and England as a whole, to provide new evidence to assist school staff, policy makers and academics in understanding the multi-faceted nature of school effectiveness and the need to evaluate school performance in detail. The objectives were to investigate: 1] the specification of the optimal models for measuring secondary school effectiveness across a range of outcomes and regions in the UK and abroad. This involved examining the size, extent and internal consistency (across outcomes, pupil groups, cohorts and curriculum stages) of schools' effectiveness and also the impact of other hierarchical levels in the education system (classrooms and primary schools) on secondary schools' effectiveness. 2] the extent of regional differences in the results. 3] the definition of the underlying dimension(s) of school effectiveness across different regional and policy contexts. The findings show that at least four dimensions of secondary school effectiveness can be defined, specifically in terms of different: outcomes (eg GCSE or attitudinal); pupil groups (Eg high or low attainers); pupil cohorts; curriculum stages (eg key stages 3 & 4). In addition regional differences appear to exist in terms of the size and impact of school effects, and these are mirrored by differences in regional context. In conclusion it is argued that effectiveness at different levels of the education system (eg individual pupils; classrooms; departments; whole school; LEA and nationally) as well as interactions between levels needs to be continually monitored in order to inform policy development and map out the *boundaries* of school effectiveness. The findings are discussed in relation to developing a value added framework for school evaluation in the UK.

Introduction

School effectiveness and the related areas of school improvement and evaluation have been topics for an increasing body of academic study since the 1960's (Coleman *et al* 1966, Jenks *et al* 1972, Rutter *et al* 1979, Mortimore *et al* 1988, Reynolds *et al*, 1996a, Hopkins, Ainscow & West 1994, Sammons, Thomas & Mortimore 1997, Gray *et al* 1999). In contrast, only in the last decade or so have policy makers focused their attention on the possibilities for improving educational practice and pupil performance via more systematic approaches to evaluation and accountability (Reynolds *et al* 1996b). Teachers and Local Education Authorities (LEAs) are now required to use performance data to inform their own evaluations of the education they provide (DfEE, 1996, 1998a; OFSTED, 1998). At the national level, the Office for Standards in Education (OFSTED) inspection reports and school league tables continue to be published as a mechanism for educational accountability. In 1998, the Department for Education and Employment (DfEE) completed a pilot 'value added' study as a supplement to league tables (SCAA 1997; DfEE 1998b).

However, in part, these policy developments have been informed by a relatively small number of quantitative research studies, often employing limited or incomplete datasets. For example, there is little research on value added secondary school effects across different pupil outcomes and regional contexts. Previous work has focused mainly on academic outcomes for a specific curriculum stage (Thomas *et al* 1997a) and few studies have looked at comparisons between regions which vary in terms of both educational policy, socio-economic and other regional factors (Scheerens & Bosker, 1997 Creemers *et al*, 1994). Also very few studies have investigated methodological developments of the value added approach such as the influence of primary schools on secondary school performance (Goldstein & Sammons 1997) or the extent of differences in effectiveness

between classes (Rowe & Hill 1994, Hill & Goldstein 1998). There is a need to develop this area of research in the UK and to clarify the findings of school effectiveness studies in the wider regional and national context using appropriate multilevel techniques. Moreover, recent DfEE and OFSTED reviews of educational research have emphasised that few quantitative studies are reported and recommend that further research is required that replicates and builds on the findings of previous work (Tooley & Darby 1998, Hillage *et al* 1998).

Objectives

The study aims to replicate, clarify and extend previous research concerning the definition and measurement of secondary school effectiveness in the UK (using multilevel techniques) across a range of outcomes and regional contexts. The objectives were:

(1) To establish the optimal multilevel model(s) for measuring school effectiveness over time using a value added approach in a range of different pupil outcomes (academic and attitudinal[1]). **Addressed by research questions 1-4.**

(2) To compare the optimal model(s) across different regional contexts (inner city, county LEAs) and education systems within the UK (England, Scotland) and also abroad (Netherlands[2]). **Addressed by research question 5.**

(3) To identify and define the dimension(s) of school effectiveness that encompass a range of different outcomes and take into account different regional, socio-economic and educational policy contexts. **Addressed by research question 6.**

Six research questions have been addressed and these are described under **Methods** and **Results**. The implications of the findings are discussed in relation to secondary school evaluation in the UK.

Methods

Research Design

The study involved a comparison of secondary school effects drawn from a variety of geographical regions. The overall focus of the analyses is on pupil outcomes at the end of statutory schooling (eg GCSE and attitudes at key stage 4), however, additional analyses also examine pupil outcomes at Key stage 3 and post 16.

The methodology adopted a 'value added' approach which adjusts for 'intake' and aims to separate and measure the school effect and that of other external factors (such pupil prior attainment and socio-economic status) on pupil performance. This approach is well established and further details of the rationale can be found elsewhere (see Nuttall *et al* 1989, McPherson 1992, Fitz-Gibbon 1995, Mortimore, Sammons & Thomas 1994, Thomas, Pan & Goldstein 1994, Sammons, Mortimore & Thomas 1996, Thomas & Mortimore 1996, Goldstein *et al* 1993, Goldstein 1997, Gray 1993,

---

[1] Originally it was intended to also examine vocational outcomes in this study. However, due to the difficulty of obtaining a representative sample of pupil's vocational outcomes (such as all GVNQ entries not just those awarded), this aspect of the project will only be discussed in terms of possible future research.

[2] It was intended that access to additional comparable datasets from abroad (eg Australia) would be agreed for inclusion in the study. However, only the data providers in the Netherlands data gave the necessary permission.

Saunders, 1998). However, it is important to acknowledge at this point that it is impossible to provide statistical adjustment for *all* factors outside the control of the school which have a significant impact on pupil performance. One limitation of the study is that the outcome and explanatory data employed were pre-selected by the data providers, although these measures have been shown to be important in previous research (Sammons *et al* 1994). Therefore, the value added effects represent the school effect *and other effects not accounted for in the analysis*. The aim is to identify the optimal multilevel model given the data available.

Samples and Data

Six datasets relating to a variety of regions in the UK and abroad (Lancashire, London, Jersey, Scotland, Netherlands, England) were employed for the study (see **Appendix 1**). Where available the datasets include individual pupil outcomes in different areas (academic and attitudinal), and results for different cohorts and curriculum stages. The following categories of explanatory variables are also included: prior attainment (or attitude) data, background factors (eg entitlement to free school meals [FSM - a measure of low family income], ethnicity, gender and age) and school context (percentage of low attaining pupils - drawn from approximately the bottom 25% ability band).

Statistical Analysis

To establish the optimal model(s) of secondary school effectiveness (**objective 1**) the statistical technique of multilevel modelling has been employed (Paterson & Goldstein 1991, Rasbash & Woodhouse 1995, 1998). The optimal model is identified in terms of purpose (ie intended use of results), statistical criteria (ie goodness of fit, statistical significance) and appropriateness (ie employing valid outcome and explanatory variables).

The analyses have been carried out in two stages (see **Appendix 2**). The first stage involved identifying which explanatory variables should be included in the optimal multilevel model(s) over time. The second stage involved extending the optimal model(s) by employing a fixed set of explanatory variables and different model specifications to examine the school residuals for different outcomes, groups of pupils, cohorts, and curriculum stages. The impact on school effectiveness of controlling for effects at other levels of the education system (classroom, region or previous school) was also examined.

Where data were available the analyses were repeated for each regional dataset and the results compared (**objective 2**). By drawing together the findings from different datasets the aim was to identify and define the *underlying* dimension(s) of school effectiveness (**objective 3**).

Results

1 Which explanatory variables should be controlled for in the optimal multilevel model(s)?

Academic Outcomes

Overall the findings suggest that in terms of statistical criteria and appropriateness the explanatory variables included in the optimal model may vary slightly for different outcomes as well as different regions (eg school context may or may not be a significant factor). However, for the purpose of employing a consistent set of explanatory variables across different regional datasets and subject outcomes - as in this study - the optimal model takes into account all prior attainment, background and school context factors. In fact, on average, this model explains the highest percentage of total (52.7%) and school (77.3%) level variance in pupil outcomes across datasets in comparison to all other models tested (**see Appendix 2**).

After controlling for all explanatory factors there are still clear differences in the extent of variation across schools for different outcomes. For example, for multiple cohort datasets, the average percentage of variance in pupil's total score outcomes attributable to schools is 7.6% (2.1% is attributable to differences between cohorts). Using considerable larger datasets, these findings replicate and confirm the results of previous studies. However, the extent of remaining differences between schools varies across regions and this issue is addressed by research question 4. In terms of educational policy these results provide strong evidence of the impact of schools and teachers on pupils' academic outcomes and illustrates the need to provide schools with feedback data on their value added performance.

Attitude Outcomes

In contrast to the multilevel results for academic outcomes, the data from one pupil cohort (1997) in Scotland shows that secondary pupils' previous attitudes, background characteristics and school context were not particularly good in explaining their later attitudes[3] (on average, the total and school percentages of variance explained were 19% and 55%). Nevertheless, in terms of statistical criteria, the optimal model for all attitude outcomes included pupils' previous attitudes and background characteristics. Further details of the analyses and findings are reported in elsewhere (Thomas *et al* 1999).

Overall the findings indicate that differences between schools in pupils' attitudes are small in comparison to the results for academic outcomes (less than 5% of the total variance in both raw and value added attitude scores is attributable to school). Nevertheless, for policy-makers and practitioners, the findings indicate the kind of affective outcomes that may be most useful to secondary schools for the purpose of providing self evaluation feedback. For example, measures that reflect pupil liking for school (ie engagement scale) and the positive interaction between teachers and pupils in the classroom (ie teacher support scale). However, further research is needed to confirm the results using data from more than one cohort and this work is continuing in collaboration with Lancashire LEA.

2. What outcomes should be employed in the optimal multilevel model(s)?

Both separate and multivariate approaches were employed to examine school effects in different academic outcomes (eg total score, language, mathematics, science). Replicating and extending previous research (eg Thomas *et al* 1997a) the results show that schools can have quite different effects in different departments and point to the existence of an effectiveness dimension for each academic subject[4]. Moreover, the consistency of schools departmental effectiveness across datasets can vary suggesting that whole school policies may have a greater impact in some regions (eg Lancashire) than in others (eg London).

The correlations between schools' adjusted residuals for academic and attitudinal outcomes indicate that the relationship between schools performance in these two areas is relatively weak (range in r: -0.38-0.19, Scotland; -0.31-0.11, Lancashire). However, as may be expected, the 'raw' unadjusted pupil level correlations are somewhat stronger (range in r: -0.27-0.61, Scotland; -0.06-0.38, Lancashire). Educationally important, these new findings support earlier work at the primary level

---

[3] ie using the attitude scales: Engagement, Pupil Culture, Self Efficacy, Behaviour, Teacher Support

[4] For example, across datasets the correlations (r) between school residuals from separate subject analyses range as follows: 0.19-0.56 (London); 0.57-0.75 (Lancashire); -0.05-0.66 (Jersey); 0.39-0.83 (Scotland); 0.44 (Netherlands).

(Mortimore *et al*, 1988) and tentatively suggest that separate dimensions of effectiveness can be identified reflecting different aspects of how schools and teachers can influence pupils' attitudes and achievements.

3. What evidence is there to suggest that the optimal multilevel model(s) should be extended to reflect school effectiveness for different pupil groups, cohorts or curriculum stages?

The optimal models were extended to identify, if possible, differential effectiveness for different pupil groups. However, this approach is only employed for academic outcomes as insufficient data were available to extend the attitude outcome models. The rationale was to examine the consistency of schools overall and departmental residuals for different pupil groups (categorised by individual background factors, cohorts and curriculum stages) in order to identify whether different effectiveness dimensions exist.

(i) Individual background factors

Replicating and extending the findings of previous work (Thomas *et al* 1997b) the correlations between school residuals for different pupil groups (categorised by prior attainment, gender and FSM) indicate that some differences exist within schools in terms of school and departmental effects for particular groups (**see Appendix 3**). The results show that non perfect correlations have been found for pupil groups categorised by prior attainment and FSM across four regional datasets with average correlations (r) of 0.59 (prior attainment) and 0.91 (FSM). In contrast, the evidence for differential effects according to gender is weak with average correlations (r) of 0.97. Interestingly, the consistency of school and departmental effects for different pupil groups categorised by prior attainment appears to be stronger in some regions (eg the Netherlands) than in others (eg London).

To examine the FSM differential results in more detail a further model has been employed which includes only prior attainment explanatory variables and therefore does not make the assumption that average attainment differences exist between particular groups. Using this approach, for example in Lancashire, only 22 percent of schools obtain positive value added scores, on average, for pupils entitled to free school meals (FSM), whereas 72 percent of schools obtain positive scores for non FSM pupils. Moreover, in nearly all schools (93%) FSM pupils make less progress on average than other pupils. In spite of the crudeness of the FSM indicator, these findings could be usefully interpreted as most schools having different levels of effectiveness for pupils who are more (or less) advantaged economically (and possibly also more/less effective learners). Nevertheless, some schools do appear to be able to decrease the attainment gap between more and less advantaged pupils. For the purpose of monitoring equal opportunities and pupil entitlement this evidence points to schools need for feedback that makes explicit the absolute levels of *progress* made by different pupil groups.

(ii) Cohorts

Four multilevel methods of identifying trends over time were examined using the London and Lancashire datasets (**see Appendix 4**).

For each method the correlations between school residuals for consecutive pupil cohorts were calculated to identify differences in school and departmental effects over time. It was found that the stability or instability of school residuals for individual cohorts varied depending on which analysis method was employed. Therefore, different methods are appropriate according to the intended purpose and use of the results. For example, using a separate analysis for each cohort (method 1) or

separate intercepts for each cohort in a joint analysis (method 2) emphasises instability over time and these methods are appropriate for the purpose of examining in detail the improvement (or decline) in value added scores over time. In contrast value added results that reflect linear trends (method 3) or the average results of two or more consecutive cohorts (method 4) are more stable over time (ie do not fluctuate randomly from year to year) and are appropriate to examine long term patterns of school performance. Indeed, method (4) has been usefully employed by schools in Lancashire LEA since 1996 as a kind of 'rolling average' of school performance (Thomas, 1998).

Interestingly, comparing the results in Lancashire (over 5 cohorts) and London (over 3 cohorts) to equivalent work by Gray, Goldstein & Jesson (1996) shows that differences between schools in value added time trends are either not statistically significant or much smaller than previously reported (having already accounted for average time trends across schools). That is, irrespective of schools' apparent improvement in raw league table performance, few schools are able to improve substantially in their effectiveness - *relative to that of other schools.*

Additional models were also employed to examine whether average trends over time in school effectiveness results varied for different groups of pupils. The findings indicate that the average time trends for particular groups (according to prior attainment and gender) appear to be slightly different. Thus further new evidence is provided that schools need to monitor equal opportunities over time.

(iii) Curriculum Stages

New and previously unreported correlations between schools' effectiveness estimates for key stage 3 and key stages 3 and 4 combined were calculated using 1997 Lancashire data to identify differences in school and departmental effects for different curriculum stages. The non perfect correlations (r = 0.50, total score; 0.27, English; 0.65, mathematics; 0.63, science) show that some schools can obtain quite different value added scores according to whether the whole or only part of the secondary curriculum is examined. The results suggest the existence of separate effectiveness dimensions for different National Curriculum stages, particularly for English outcomes. For government policy makers these findings are important given the publication of a sample of schools value added results for key stage 4 only (DfEE, 1998b). In the light of the current findings a school could appear to be doing well at key stage 4, but not so well at key stage 3 or across both key stages 3 and 4. Overall the results indicate that separate value added measures of effectiveness should be feed back to schools for each key stage as well as for the whole period of secondary schooling.

4. What evidence is there to suggest that the optimal multilevel model(s) should be extended to incorporate additional hierarchical groupings (or levels) within the education system (ie classrooms, regions or previous school)?

Following on from the analyses addressing research question 3 the optimal models for each academic outcome measure were again extended to establish the need to take account of effects attributable to levels other than the school level within the education system. The significance of other levels in accounting for pupil attainment and relative progress is crucial evidence in identifying more precisely the dimensions and appropriate methods of calculating school effects.

(i) Previous Schools

Using 1997 Lancashire data the value added results from cross-classified multilevel models (Goldstein 1995) show that the total and school level variance as well as the variation in pupil outcomes attributable to secondary schools is only slightly reduced (or similar) when allowance is made for primary school attended. Primary school attended does have a statistically significant (at

0.05 level) impact on pupils' GCSE outcomes, but the total variance in value added scores attributed to the primary level is small (ranging from 1%-3%) compared to the secondary level (ranging from 6-12%). These results are in contrast to previous similar work in the UK by Goldstein & Sammons (1997) which reports larger primary school effects than secondary school effects. Interestingly, the findings are in line with similar results in the Netherlands (Snijders & Bosker, 1999) and this suggests that the Goldstein findings may not be representative of the general pattern in UK schools due possibly to the small and incomplete sample employed.

The correlations between school residuals with and without controls for primary school attended are very high (r> = 0.99). This indicates that, for the practical purpose of secondary school feedback, little difference in interpretation may be drawn from value added measures which do, or do not, make allowance for primary school attended. Nevertheless, for the purpose identifying overall patterns in and underlying influences on secondary pupils' attainment further work using cross-classified models is required to examine the impact of primary school effectiveness on later achievements by linking *progress* made by pupils at the primary level to subsequent *progress* at secondary school. This approach is also required to examine the related issue of pupils moving school *within* the primary or secondary phase of schooling.

(ii) Classrooms

Using English and mathematics outcome data from the 1997 Scottish dataset the value added results show that total and school level variance as well as the variation in pupil outcomes attributable to secondary school is only slightly reduced when allowance is made for variation between classes[5] as well as between schools. Nevertheless, some statistically significant (at 0.05 level) variation in pupil performance can be attributed to differences between classes (or teachers) and the impact is considerably greater for mathematics than English outcomes. For English the impact of classroom differences on the estimates of school effects is small for both raw outcomes and in terms of value added (percentage of variance attributable to the class level is 2.2% and 1.8% respectively). In contrast, for mathematics, the variation in raw outcomes attributed to class differences is larger (23.7%) and comparable to previous research where classroom effects on mathematics have been estimated for Scotland (Scheerens, Vermeulen & Pelgrum 1989, Scheerens & Bosker 1997). However, the variation in mathematics outcomes attributed to classes is reduced considerably for value added measures (5.3%). These findings point to the influence of setting in classroom organisation which is more prevalent in mathematics than English teaching. Nevertheless, for mathematics in particular, these results suggest that within school differences in terms of classroom (or teacher) effectiveness may need to be addressed.

Similar to the results adjusting for primary school level, the correlations between school residuals with and without controls for classroom level are very high (r> = 0.99). Interestingly, these results suggest that for the practical purpose of providing departmental (or subject) feedback little difference in interpretation may be drawn from value added estimates which do, or do not, make allowance for classroom grouping - at least in Scotland. However, further research is required with larger datasets and regions to tease apart the impact on pupil attainment of classroom (or teacher) effectiveness and other school or contextual factors (such as setting, classroom organisation, or teacher mobility).

---

[5] It is important to note that the variation in pupil progress is estimated in relation to classroom grouping in 1995. Therefore, for the purpose of this analysis it is assumed that few pupils change classes during the two year period examined 1995-97. In cases where this assumption is incorrect, examining class effects over a one year period or using cross-classified models may be more appropriate.

(iii) Regions

Employing small area definitions of region (involving approximately 10 or fewer secondary schools) no statistically significant regional effects were identified in the Lancashire or London datasets. However, findings from an additional dataset comprising the 1994-95 A/As level results for the whole of England was used to examine the impact of regional effects categorised by LEA. The results show that at the post 16 level small differences in effectiveness between LEAs do appear to exist (percentage of variance attributable to the region level was less than 3% for all outcomes tested). It is unclear whether this conclusion would change for secondary results at key stages 3 and 4, although a replication of this analysis will be possible in the future once national curriculum assessments become available nationally.

Importantly, the above results use a single model for England as a whole which controls for the average differences in pupil attainment (according to prior attainment and other factors) across all regions. In research question 4 a meta analysis approach is employed by examining the regional differences via the separate model results for each regional dataset (*cf.* Bosker & Witziers, 1995).

5. Are different optimal multilevel model(s) required to reflect regional differences?

The evidence reported for research questions 1-4 show there is some variability in results across regional datasets. These findings tentatively suggest that regional context or policy does have an influence on the range and extent of school performance and points to the value of separate regional measures of school performance. However, some differences observed may be due to differences in the availability or definition of explanatory variables included in the optimal models. Therefore, in order to provide further evidence on regional differences the results of models which employ only pupils' prior language attainment have also been examined.

The findings show that regions vary in terms of the goodness of fit of models using only language prior attainment as an explanatory factor. However, some regions are more alike than others with Lancashire, Jersey and Scotland obtaining fairly similar results and a much better fit of the data (average percentage total variance explained ranges from 41% to 45%) in contrast to the equivalent results for London and the Netherlands (24%-32%). Moreover, the average percentage of school level variance explained varies across regions (from 51% to 68%). These differences are larger than would be expected on the basis of differences in the method of language assessment[6] between regions and suggest that certain aspects of regional policy or context may be linked to a weaker relationship between pupils' prior and later attainment. For example, London as an inner city area has additional factors of pupil and community poverty as well as long established equal opportunity policies that may result in pupils' outcomes not being predicted particularly well on the basis of their previous reading ability.

The size of school effects also varies considerably between regions. However, as noted above, some regions obtain more similar results than others. Lancashire, London and Scotland - interestingly all within the UK - obtain similar variations between schools in terms of effectiveness with differences of approximately 0.8 standard score units between the most and least effective schools[7]. In contrast, Jersey and the Netherlands appear to have much larger variations in schools' effectiveness with pupil outcome differences of more than 1.3 standard score units between the most and least effective

---

[6] The method of assessment varied in some cases (eg multiple choice, reading comprehension).

[7] ie calculated using +/- 2 standard deviations of schools' value added residuals.

schools. The percentage of total variance in value added scores attributable to schools mirrors the regional differences in the size of school effects. Regions again appear to fall into two groups, the first group (Lancashire, London and Scotland) obtain results showing 6-9% of the variance in pupil outcomes is attributable to schools. In comparison, the second group (Jersey, the Netherlands) show a much greater impact of schools (22-24% of variance attributable to schools).

In order to illuminate these findings contextual information about each region has also been examined in terms of (I) the extent of selection amongst schools (ie standard deviation across schools of the percentage of low attaining pupils in each school) and (ii) school differences in raw unadjusted outcomes (ie standard deviation across schools of raw residuals). Interestingly, the regional differences in terms of context reflect the differences observed in terms of the size and impact of school effects. It appears that regions with larger differences between schools in raw scores and the extent of selection (Jersey, the Netherlands[8]) also show larger differences in the size and impact of school effects.

These findings provide further evidence of regional differences in school effects. Therefore, it seems likely that regional, as well as national, context and policies do have an influence on schools' effectiveness, although a causal link has not been tested. In other words regional context (such as socio-economic or geographical facors) or education policy (such as the extent of selection or private schooling) may limit the possibilities of a school being more or less effective as well as enhance (or inhibit) the overall differences between schools. Moreover, regional differences in context or policy may also be reflected in how well pupil's previous attainment can predict their later attainment. The results highlight the need for further analyses of the size and impact of schools effects in different UK regions, particularly those that vary greatly in terms of selection policy (such as LEAs with a system of selective grammar schools).

6. Does the evidence suggests different dimension(s) of secondary school effectiveness exist and if so how should these be defined?

The findings from research questions 1-3 suggest that at least four dimensions of secondary school effectiveness can be defined specifically in terms of different outcomes, pupil groups, pupil cohorts and curriculum stages. To clarify these definitions the impact or constraints on schools' effectiveness associated with other levels within the education system have also been examined. The findings from research question 4 tentatively suggest that pupils' attendance at a particular primary school, or the classroom grouping of pupils (or at the post 16 level, the location of the institution in a particular LEA), has a relatively small impact in practical terms on the measurement of how effective a school is.

Nevertheless, effectiveness at different levels of the education system (eg individual pupils; classrooms; departments; whole school; regionally and nationally) as well as the interaction between levels need to be continually monitored in order to inform policy development and map out the *boundaries* of school effectiveness. The evidence from research question 5 of regional differences in the range and extent of school effects points to the *interpretation* of school effects being regionally dependent and highlights the important role of regional or policy context when defining or evaluating school performance. For example, in England, separate LEA analyses of school performance may be needed in addition to the national level.

---

[8] The Netherlands employs a system of selecting pupils into different academic and vocational secondary schools.

These findings are of theoretical as well as practical importance pointing to the possible existence of regional sub-systems in the overall functioning of schooling (*cf.* Scheerens & Bosker, 1997). However, further work is needed to investigate and explain the relationship between specific aspects of regional context or policy and the average levels as well as the range of schools' effectiveness in different regions. This approach is in line with other researchers who have emphasised the importance of examining in detail the socio-economic context of pupils and schools in relation to their effectiveness (Mortimore & Whitty 1997, Thrupp 1997, Lauder, Jamieson & Wikeley 1998).

Conclusions

Replicating and extending previous work, the results of this study emphasise the need for school staff and external evaluators to analyse value added data in a sensitive and detailed way. To measure schools' 'value added' pupil attainment measures are needed at the beginning and end of each curriculum stage[9] as well as other background data. A valid framework for secondary school evaluation in the UK needs to incorporate at least four *underlying* dimensions of effectiveness (in terms of different outcomes, pupil groups, cohorts and curriculum stages) and also needs to contextualise the results with regional information. To operationalise this kind of evaluation the data, methods and their limitations need to be transparent and well understood by the evaluators (Goldstein & Myers, 1996).

Hopefully, evidence of this kind will stimulate and inform teachers' evaluation of their own educational practices and capacity for improvement as well as make transparent the constraints, boundaries and context of schools' effectiveness. However, the findings also point to the need for further research on the existence of additional dimensions of school effectiveness in terms of vocational and other pupil outcomes valued by society. Indeed, a comprehensive value added framework for school evaluation might also encompass measures related to numerous other aspects of a school's goals, processes and outcomes (MacBeath & Mortimore 1994, Thomas *et al* 1998).

Further Research Priorities

Research examining secondary school effects in terms of a wider range of outcomes is required as little data is currently available to address this issue. In particular, more detailed measures of pupils' affective and vocational outcomes need to be developed and analysed to identify the influence of schooling in these important areas. This work is crucial given the considerable emphasis currently given to schools' academic performance in response to the publication of national league tables. Moreover, there is a need for detailed case studies (in a range of regional, policy or socio-economic contexts) to describe the long-term impact of accountability and school self-evaluation processes on the quality of teaching and learning. Research is also required to extend the datasets currently available and develop further the methodology of analysing secondary school effects. In the broader context, the application of detailed value added evaluation methods also need to be explored for primary, post 16, further and higher education.

---

[9] It should be noted that national prior attainment measures for pupils entering secondary school/key stage 3 are not currently available.

References

Bosker, R & Witziers, B (1995) *School Effects, Problems, Solutions and a Meta-Analysis*, paper presented at the International Congress for School Effectiveness & School Improvement, Leeuwarden, The Netherlands, January 1995.

Coleman, J S, Campbell, E, Hobson, E, McPartland, J, Mood, A, Weinfeld, F & York, R (1966) *Equality of Educational Opportunity*, Washington: National Center for Educational Statistics.

Creemers, B P M, Reynolds, D & Swint, F E (1994) *First Results of the Quantitative Study*, paper presented at the British Education Research Association, Oxford, September 1994.

DfEE (1996) *Setting Targets to Raise Standards: A Survey of Good Practice*, London: Department for Education and Employment.

DfEE (1998a) *The Autumn Package: Pupil Performance Information*, London: Department for Education and Employment.

DfEE (1998b) *1998 Value Added Pilot: Supplement to the Secondary School Performance Tables*, London: Department for Education and Employment.

Fitz-Gibbon, C (1995) *National Value Added Project. General Report: Issues to be Considered in the Design of a National Value Added System*, London: School Curriculum and Assessment Authority.

Goldstein, H (1995) *Multilevel Statistical Models (2nd Edition)*, London: Edward Arnold & New York: Halsted Press.

Goldstein, H (1997) Methods in School Effectiveness Research, *School Effectiveness & School Improvement*, 8, (4): 369-395.

Goldstein, H, Rasbash, J, Yang, M, Woodhouse, G, Pan, H, Nuttall, D & Thomas, S (1993) A Multilevel Analysis of School Examination Results, *Oxford Review of Education*, 19, (4): 425-433.

Goldstein, H & Myers, K (1996) Freedom of Information: Towards a Code of Ethics for Performance Indicators, *British Educational Research Association Newsletter: Research Intelligence*, July, 12-16.

Goldstein, H & Sammons, P (1997) The Influence of Secondary and Junior Schools on Sixteen Year Examination Performance: A Cross-Classified Multilevel Analysis, *School Effectiveness & School Improvement*, 8, (2): 219-230.

Gray, J (1993) *Value-Added Approaches in School Evaluation: The Experiences of Three LEAs in England - Lessons and Challenges*, Audit Unit, HM Inspectors of Schools, The Scottish Office.

Gray, J, Goldstein, H & Jesson, D (1996) Changes and Improvements in Schools' Effectiveness: Trends over five years, *Research Papers in Education*, 11, (1): 35-51.

Gray, J, Hopkins, D, Reynolds, D, Wilcox, B, Farrell, S & Jesson D (1999) *Improving Schools: Performance and Potential*. Buckingham: Open University Press, ISBN 0-335-20398-1.

Hill, P & Goldstein H (1998) Multilevel Modelling of Educational Data With Cross–Classification and Missing Identification for Units, *Journal of Educational and Behavioural Statistics*, 23, (2): 117-128.

Hillage, J, Pearson, R, Anderson, A & Tamkin, P (1998) *Excellence in Research on Schools*, London: Department for Education and Employment.

Hopkins, D, Ainscow, M & West, M (1994) *School Improvement in an Era of Change*, London: Cassell.

Jencks, C S, Smith, M, Ackland, H, Bane, M J, Cohen, D, Gintis, H, Heyns, B & Micholson, S (1972) *Inequality: A Reassessment of the Effect of Family and Schooling in America*, New York: Basic Books.

Joreskog, K G & Sorbom, D (1989) *LISREL 7: User's Reference Guide*, Chicago: SPSS Inc. (Appendix)

Lauder, H, Jamieson, I & Wikeley, F (1998) Models of Effective Schools: Limits and Capabilities, in R Slee, G Weiner & S Tomlinson (Eds) *School Effectiveness for Whom?*, London: Falmer.

MacBeath, J & Mortimore, P (1994) *Improving School Effectiveness – A Scottish Approach*, British Educational Research Association 20th Annual Conference, September, St Anne's College, Oxford.

McPherson, A (1992) Measuring added value in schools *National Commission on Education Briefing No 1*, February 1992, London.

Madaus, G F, Kellaghan, T, Rakow, E A & King, D (1979) The sensitivity of measures of school effectiveness, *Harvard Educational Review*, 49, 207-230. (Appendix)

Mortimore, P, Sammons, P, Stoll, L, Lewis, D & Ecob, R (1988) *School Matters: The Junior Years*, Wells: Open Books.

Mortimore, P, Sammons, P & Thomas, S (1994) School Effectiveness and Value Added Measures, *Assessment in Education: Principles, Policy & Practice*, 1, (3): 315-332.

Mortimore, P & Whitty, G (1997) *Can Improvement Overcome the Effects of Disadvantage?*, London: Institute of Education, 16.

Nuttall, D, Goldstein, H, Prosser, R & Rasbash, J (1989) Differential School Effectiveness, Chapter 6 in *International Journal of Educational Research*, special issue Developments in School Effectiveness Research, 13, 769-776.

OFSTED (1998) *School Evaluation Matters*, London: OFSTED

Paterson, L & Goldstein H (1991) New statistical methods of analysing social structures: an introduction to multilevel models, *British Educational Research Journal*, 17, (4): 387-393.

Rasbash, J & Woodhouse, G (1995) *Users' Guide for MLN*, London: Institute of Education, University of London.

Rasbash, J & Woodhouse, G (1998) *Users' Guide for MLwiN*, London: Institute of Education, University of London.

Reynolds, D, Bollen, R, Creemers, B, Hopkins, D, Stoll, L & Lagerweij, N (Eds) (1996a) *Making Good Schools: Linking School Effectiveness and School Improvement*, London: Routledge.

Reynolds, D, Sammons, P, Stoll, L, Barber, M & Hillman, J (1996b) School Effectiveness and School Improvement in the United Kingdom, *School Effectiveness & School Improvement (special issue of country reports)*, 7, (2): 133-158.

Rowe, K J & Hill, P W (1994) *Multilevel Modelling in School Effectiveness Research: How Many Levels?* paper presented at the International Congress for School Effectiveness & Improvement, Melbourne, Australia, 3-6 January 1994.

Rutter, M, Maughan, B, Mortimore, P & Ouston, J (1979) *Fifteen Thousand Hours: Secondary Schools and Their Effects on Children*, London: Open Books.

SCAA (1997) *The Value Added National Project - Report to the Secretary of State*, London: SCAA.

Sammons, P, Thomas, S, Mortimore, P, Owen, C & Pennell, H (1994) *Assessing School Effectiveness: Developing Measures to put School Performance in Context*, London: Office for Standards in Education [OFSTED].

Sammons, P, Mortimore, P & Thomas, S (1996) Do schools perform consistently across outcomes and areas? To appear in J Gray, D Reynolds, C Fitz-Gibbon & D Jesson (Eds) *Merging Traditions: The future of research on school effectiveness and school improvement*, London: Cassells.

Sammons, P, Thomas, S & Mortimore, P (1997) *Forging Links: Effective Schools and Effective Departments*, London: Paul Chapman, ISBN 1-85396-349-6.

Saunders, L (1998) *'Value Added' Measurement of School Effectiveness: An Overview*, Slough: National Foundation for Education Research.

Scheerens, J, Vermeulen, C & Pelgrum, W (1989) Generalizability of instructional and school effectiveness indicators across nations, *International Journal of Educational Research*, 13, 789-800.

Scheerens, J & Bosker, R (1997) *The Foundations of Educational Effectiveness*, Oxford: Elsevier Science.

Snijders, T & Bosker, R (1999) *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, London: Sage.

Thomas, S (1998) Value Added Measures of School Effectiveness in the United Kingdom, *Prospects*, 28, (1): 91-108.

Thomas, S, Pan, H & Goldstein, H (1994) *Report on Analysis of 1992 Examination Results* prepared for the AMA, London: Institute of Education, University of London.

Thomas, S & Mortimore, P (1996) Comparison of Value Added Models for Secondary School Effectiveness, *Research Papers in Education* , 11, (1): 5-33.

Thomas, S, Sammons, P, Mortimore, P & Smees, R (1997a) Stability and Consistency in Secondary Schools' Effects on Students' GCSE Outcomes over Three Years, *School Effectiveness and School Improvement*, 8, (2): 169-197.

Thomas, S, Sammons, P, Mortimore, P & Smees, R (1997b) Differential Secondary School Effectiveness: Examining the size, extent and consistency of school and departmental effects on GCSE outcomes for different groups of students over three years, *British Educational Research Journal*, 23, (4): 451-469.

Thomas, S, Smees, R, MacBeath, J, Sammons, P & Robertson, P (1998) *Creating a Value Added Framework for Scottish Schools*, report prepared for the Scottish Office Education & Industry Department, Edinburgh.

Thomas, S & Smees, R (1998) *Dimensions of Secondary School Effectiveness: Comparing the findings from four academic studies*, paper presented at the annual conference of the American Educational Research Association, San Diego, 13-17 April.

Thomas, S, Smees, R, MacBeath, J, Robertson, P & Boyd, B (1999) *Valuing Pupils' Views in Scottish Schools (*to appear in *Educational Research & Evaluation*), London: Institute of Education, University of London.

Tooley, J & Darby, D (1998) *Educational Research: A Critique*, London: OFSTED.

Thrupp, M (1997) *The School Mix Effect: How the Social Class Composition of School Intakes Shapes School Processes and Student Achievement*, paper presented at the Annual Meeting of the American Education Research Association, Chicago, 24-28 March.

Appendix 1 : Samples from Six Datasets

(1)	**LANCASHIRE [1993-97] 99 Schools; 61,103 students**	
	Outcomes:	*GCSE scores (total score, English, mathematics, science) Key Stage 3 levels (total score, English, mathematics, science) and attitude scales (Engagement, Pupil Culture, Self Efficacy, Behaviour, Teacher Support).*
	Baselines:	*Prior attainment (verbal, quantitative, non-verbal[10]) at age 11 (year 7).*
	Background:	*gender, age, entitlement to free school meals, ethnicity.*
	Context:	*% low attainers.*
(2)	**LONDON [1990-92] 94 schools; 17,850 students[11]**	
	Outcomes:	*GCSE scores (total score, English, mathematics, science).*
	Baselines:	*Prior attainment (London reading test, Ability bands) at age 11 (year 7).*
	Background:	*gender, age, entitlement to free school meals, ethnicity.*
	Context:	*% low attainers.*
(3)	**JERSEY [1993-95] 9 schools; 1,849 students[12]**	
	Outcomes:	*GCSE scores (total score, English, mathematics, science).*
	Baselines:	*Prior attainment (Reading test, verbal and non-verbal tests) at age 11 (year 7).*
	Background:	*gender, age.*
	Context:	*% low attainers.*
(4)	**SCOTLAND [1997] 36 schools; 4,500 students**	
	Outcomes:	*standard grade examinations (total score, English, mathematics, science) and attitude scales(Engagement, Pupil Culture, Self Efficacy, Behaviour, Teacher Support).*
	Baselines:	*Prior attainment (reading and mathematics tests) at age 13+ (S2).*
	Background:	*gender, age, entitlement to free school meals.*
	Context:	*% low attainers.*
(5)	**NETHERLANDS [1995] 256 schools; 8,543 students[13]**	
	Outcomes:	*Dutch language, mathematics attainment of 14/15 years Odds*
	Baselines:	*Prior attainment (IQ test, Dutch language, mathematics, information tests) at age 13.*
	Background:	*gender.*
	Context:	*% low attainers.*
(6)	**ENGLAND [1994-95] 2,700 Institutions; about 500,000 students[14]**	
	Outcomes:	*A/AS level examinations*
	Baselines:	*Prior attainment (GCSE scores - total, average, English, mathematics).*
	Background:	*gender, age.*
	Context:	*% low attainers.*

---

10 National Foundation of Educational Research (NFER) Cognitive Abilities Test (CAT).

11 The London dataset was previously employed in the ESRC funded study *Differential School Effectiveness: Department variations in GCSE attainment* (R000234130).

12 Sample includes all Jersey state secondary schools. However, as the sample size is small the results should be treated with caution.

13 The Netherlands dataset was originally created by Prof Jaap Scheerens and Prof Roel Bosker who gave permission for the data to be employed in this study. Only a sample of pupil records is available for each school.

14 A/AS level dataset was provided by the DfEE

Notes: Pupil records with missing prior attainment and background data were excluded from the analyses (on average missing values ranged from 4% to 36% across datasets). All pupil attainment data has been transformed to standardised scores (ie normal scores with a mean of 0 and a standard deviation of 1). All student and school background data has been recoded (where appropriate) into categories that are standard across all datasets. The five pupil attitude scales (Engagement, Pupil Culture, Self Efficacy, Behaviour, Teacher Support) were created by weighting questionnaire items using the statistical technique of Linear Structural Equation Modelling for Latent Variables (LISREL see Joreskog & Sorbom, 1989). A description of how the pupil attitude scale data have been created is provided in the nominated paper *Valuing Pupil Views in Scotland* (Thomas *et al*, 1999).

Appendix 2: Statistical Analyses

Stage (1) Identifying the Explanatory Variables

Eight different models have been contrasted in order to identify which variables should be controlled in the optimal multilevel model(s) for each regional dataset and outcome (see Panel 1). For datasets which comprise data for only a single cohort of students a two level model has been used to provide estimates of the average (fixed) effects for each explanatory variable as well as the random variation associated with the level of the school and the individual student. In the case of datasets which include more than one cohort of students (eg 1993-97 Lancashire dataset) a three level model is applied which estimates the average (fixed) effects across cohorts. Essentially, in these three level models, the school residuals represent the mean school effects over several cohorts of student (thus providing more robust estimates). In contrast, the two level models provide school residuals relating to school effects for a single cohort.

Panel 1: For each region and cognitive outcome measure eight different models have been contrasted:

Model 1	Intercepts only
Model 2	Pupil background only (ie where available: gender, age, FSM, ethnicity)
Model 3	Language prior attainment measure only
Model 4	IQ/ability prior attainment measure only
Model 5	All prior attainment measures
Model 6	Prior attainment and pupil background (ie where available: gender, age, FSM, ethnicity)
Model 7	Prior attainment measures and school context variable (ie % low attainers)
Model 8	All measures: prior attainment, background, context

An equivalent set of models were employed for the pupil attitude outcomes

Four aspects of the multilevel analyses have been examined:

[i] The weighting and statistical significance of different explanatory variables included in the models.

[ii] The percentage reduction in the total and school level variation in pupil outcomes by introducing different explanatory variables in the multilevel model as well as interactions between explanatory variables.

[iii] The percentage of total variation attributable to the pupil, cohort and school levels.

[iv] Standard deviation and range of school residuals.

The results for academic outcomes are summarised briefly below. The results for attitudinal outcomes are summarised in the nominated paper *Valuing Pupil Views in Scotland* (Thomas et al, 1999).

Academic Outcomes

The results from 144[15] separate analyses show that across all datasets and academic outcomes the impact on pupil attainment of nearly all explanatory variables tested was statistically significant (at 0.05 level). However, some differences were observed across regional datasets. For example, the results from Model 8, show all explanatory variables were statistically significant (at 0.05 level) except: IQ (Dutch [Netherlands]), Black and Caribbean ethnic groups (all outcomes [Lancashire]), gender (maths [Scotland]), age (total score, maths [London]; total score, maths, English [Scotland]), percentage low attainers (all outcomes [London], English, science [Lancashire]).

On average, the goodness of fit of the models was maximised with Model 8 (which includes all explanatory factors). The average percentage of total variance explained across all Model 8 results was 52.7% and the average school level variance explained was 77.3%. However, comparing between Models 7 and 8 only a small average improvement was achieved by including the school context variable 'percentage low attainers' (for Model 7 the average percentage of total variance explained was 50.6% and the average school variance explained was 72.3 %). In fact, for the Lancashire dataset Model 7 was very slightly better (by <0.6% total variance) for all outcomes.

Comparing between Models 3, 4 and 5, by far the largest percentage of total variance in pupil performance is explained by including all prior attainment variables (average total variance explained is 48.1%) with the language prior attainment measure being the most important (average total variance explained is 38.0%). In contrast, the IQ prior attainment measure alone (where data is available) explains less of the overall variance (27.4%). This finding supports previous evidence reporting the relative importance of employing different types of prior measures relating to the curriculum or to underlying abilities (Madaus *et al* 1979) and indicates the need to adjust for previous levels of attainment, especially first language, using appropriate curriculum based assessments in preference to IQ type tests.

Stage 2 Extending the Model

The multilevel analyses were extended using a fixed set of explanatory variables (ie those identified in *Stage 1*) and a variety of different model specifications including fitting complex variation, including additional levels [classrooms and regions], multivariate and cross-classified analyses (see Goldstein 1995,1997 for technical details) . Where data were available the analyses were repeated for each outcome and region and the results compared. The following correlational analyses were carried out to examine the relationship between school residuals from different model specifications:

[i] Correlations between school residuals for different academic outcomes (total score, language, mathematics, science) to identify differences in schools' overall and departmental

---

[15] 144 analyses = 8 models x 4 outcomes (total score, English, mathematics, science) x 4 datasets (Lancashire, London, Jersey, Scotland) + 8 models x 2 outcomes (Dutch, mathematics) x 1 dataset (the Netherlands).

performance.

[ii] Correlations between school residuals for academic (see [I]) and attitudinal (engagement with school, pupil culture, self efficacy, behaviour, teacher support) outcomes to identify differences in school performance in two areas: cognitive and affective.

[iii] Correlations between school residuals for specific pupil groups (categorised by prior attainment, gender, FSM) to identify differences in school and departmental effects for different groups.

[iv] Correlations between school residuals for different cohorts of pupils' aged 16 years (categorised by year of taking outcome assessment such as GCSE) to identify differences in school and departmental effects over time.

[v] Correlations between school residuals for different year groups (categorised, by outcome assessment at Year 11 [Key stage 4] and Year 9 [Key stage 3]) to identify differences in school and departmental effects for different curriculum stages.

[vi] Correlations between school residuals with and without adjustment for primary school attended to identify the impact of primary school attended on school and departmental effects.

[vii] Correlations between school residuals with and without adjustment for classroom grouping to identify the impact of classroom differences on school and departmental effects.

See main report for a summary of the correlational results (also Appendix 3&4).

Appendix 3: Correlations between school effects[16] for different pupil groups (Table shows range in results across subjects)

	London [1990-92] School n=94 Pupil n=17850	Lancashire [1993-97] School n=99 Pupil n=60226	Scotland [1997] School n=36 Pupil n=4406	Netherlands [1995] School n=258 Pupil n=8543
Band 1 v Band 3	0.44-0.79 [0.35-0.66]	0.60-0.86 [0.55-0.81]	0.05-0.55 [-0.14-0.39]	0.86-0.94 [0.63-0.72]
Boys v Girls	0.92-1.00 [0.81-0.92]	0.97-0.99 [0.95-0.98]	0.90-0.99 [0.81-0.94]	0.98-1.00 [0.88-0.97]
FSM v non FSM	0.66-0.94 [0.67-0.88]	0.89-0.98 [0.84-0.95]	0.94-0.99 [0.79-1.00]	n/a

Notes: [ ] = true correlations in brackets; Band 1 = top 25% in terms of prior attainment; Band 3 = bottom 25% in terms of prior attainment; FSM=free school meals; n/a= data not available.

[16] The initial complex variation models were developed further to identify if any confounding between different pupil groups would change the results. Across both datasets employed (London and Lancashire) only differential school effects by gender became not statistically significant when tested alongside prior attainment and entitlement to free school meals in the Mathematics and Science analyses. These findings provide further evidence that statistically significant within school differences in effectiveness exist for particular pupil groups, having already controlled for the average group differences in attainment.

Appendix 4: Correlations between school effects for different pupil cohorts (Table shows range in results across subjects)

	Method (1) Separate analyses (2 level)	Method (2) Joint analysis with separate intercepts (2 level)	Method (3) Joint analysis with intercept and slope (2 level[17])		Method (4) Separate 3 cohort analyses (3 level)
Consecutive pupil cohorts differing by 1 year	0.61-0.91 [0.50-0.87]	0.52-0.83 [0.33-0.70]	0.97-0.99 [0.77-0.98]	1993-5 pupil cohorts vs 1994-96 pupil cohorts	0.93-0.96
Pupil cohorts differing by 2 years	0.59-0.86 [0.27-0.85]	0.50-0.76 [0.31-0.50]	0.89-0.96 [0.39-0.91]	1993-5 pupil cohorts vs 1995-97 pupil cohorts	0.81-0.88
Pupil cohorts differing by 3 years	0.55-0.79	0.51-0.66	0.77-0.89	1994-96 pupil cohorts vs 1995-97 pupil cohorts	0.88-0.95
Pupil cohorts differing by 4 years	0.58-0.72	0.50-0.59	0.63-0.81		

Notes: [ ] = equivalent London results where available; n/a= data not available. London [1990-92]School n=94 Pupil n=17850.Lancashire[1993-97]School n=99Pupil n=60226.

c:\thomas\sesi2000.pap\22.2.00.

[17] An equivalent 3 level model was also examined and reported in the main text (*cf* Gray *et al*, 1996).